

УДК 303.432

Горбачик О.А.

Київський міжнародний інститут соціології старший науковий співробітник

СТАНДАРТ МЕТАДАНИХ ДЛЯ ФАЙЛІВ В ІНТЕРНЕТ ЯК АЛЬТЕРНАТИВА ОРГАНІЗАЦІЇ КОМП'ЮТЕРНИХ АРХІВІВ СОЦІАЛЬНИХ ДАНИХ

Стаття присвячена розгляду проблеми стандартизації метаданих для опису даних емпіричних соціологічних досліджень. Зроблений висновок про те, що стандарт метаданих у сукупності із відповідними інструментальними засобами призведе до ситуації, коли потенційно придатні для аналізу дані будуть знаходитися значною мірою поза електронними ресурсами архівів та банків соціологічних даних.

Ключові слова: метадані, емпіричне соціологічне дослідження, архів соціальних даних, стандарт опису даних, пошук в Інтернет, аналіз соціальних даних

Статья посвящена рассмотрению проблемы стандартизации метаданных для описания данных эмпирических социологических исследований. Сделан вывод о том, что стандарт метаданных в совокупности с соответствующими инструментальными средствами приведет к ситуации, когда потенциально пригодные для анализа данные будут находиться в значительной мере вне электронных ресурсов архивов и банков социологических данных.

Ключевые слова: метаданные, эмпирическое социологическое исследование, архив социальных данных, стандарт описания данных, поиск в Интернет, анализ социальных данных

Article is devoted for the problems of standards development for metadata describing empirical sociological research data. Author concludes that metadata standards with necessary and relevant information instruments will lead the situation when appropriate for analysis datafiles will rather often locate out of the resources of social databanks and social data archives.

Keywords: metadata, empirical sociological research, social data achieve, data description standard, search in the Web, social data analysis

Актуальність. В світі накопичено багато емпіричних даних, що характеризують різні соціальні об'єкти. Спектр простягається від результатів опитувань окремих респондентів до агрегованих показників перебігу певних соціальних процесів або стану справ в країнах та регіонах. Продукують ці дані органи державної статистики, академічні інститути та університети, державні або незалежні аналітичні агенції тощо. Значна частина цих даних знаходиться або у вільному доступі або принаймні є доступною дослідникам для некомерційного використання. В зв'язку із цим важливими є питання коректного використання цих даних. В світі розробляються та удосконалюються закони про авторське право, про персональні дані тощо. Актуалізуються питання професійної етики, правил використання зібраних в ході дослідження даних. Загальновизнаною є практика, коли дослідницький план кожного проекту обов'язково має бути схвалений незалежною експертною комісією (Institutional Review Board, IRB), що опікується охороною прав людей, які брали участь у дослідженні. Зокрема в Україні експертизу для соціологічних проектів здійснює Комісія з професійної етики Соціологічної асоціації України.

Все більшу роль у діяльності дослідників, що працюють у галузі соціальних наук, відіграє Інтернет. З одного боку Інтернет забезпечує середовище для професійного спілкування та спільної роботи в рамках колективів, що є розпорощеними по всьому світу. Саме "завдяки" Інтернет з'явилося поняття

віртуальної лабораторії і поширення набула дистанційна співпраця фахівців. З іншого боку сам Інтернет стає середовищем для проведення досліджень і джерелом емпіричних даних. Мова йде не тільки і не стільки про те, що традиційне для емпіричних соціологічних досліджень збирання даних переноситься у середовище Інтернет (web-опитування, фокус-групи в електронних чатах, контент-аналіз в комп'ютерних архівах газет або стрічках новин тощо), але про те, що на різних електронних сторінках знаходиться у різному вигляді (у різному форматі) багато різних даних емпіричних соціальних досліджень (масових та експертних опитувань, маркетингових досліджень, економічної та соціальної статистики тощо), що є придатними для цікавого та плідного вторинного аналізу. Мова йде про дані як кількісних так і якісних досліджень. Інтернет забезпечує дуже дешеву та ефективну інфраструктуру доступу до різноманітних за форматом та обсягом даних. В сучасних умовах зберігання даних в Інтернет (на віддалених файлових серверах або у "хмарах") є не тільки дешевим, але досить надійним та безпечним. Інколи навіть говорять про нову еру соціальних досліджень, що характеризується значним обсягом оцифрованих та практично безкоштовних емпіричних даних, які лише треба знайти, відібрати, погодити із задачами конкретного дослідження та включити до відповідного аналізу. Часто доступ до таких соціальних даних в Інтернеті є безкоштовним або дуже дешевим і єдиною справжньою перешкодою до активного використання цих даних є те, що зазвичай ці дані є слабо та нестандартно описаними. Є дані, але не вистачає "даних про дані", або як прийнято зараз говорити, не вистачає метаданих. Саме брак метаданих стає перешкодою до активного використання вже зібраних даних, до вторинного аналізу таких даних. Отже **метою** стаття є розгляд проблеми стандартизації метаданих для опису даних емпіричних соціологічних досліджень.

Вторинний аналіз використовується в дослідницькій практиці в світі все частіше. Зростає кількість публікацій, зроблених на аналізі емпіричних даних, взятих з публічно доступних архівів. Сучасні дослідження, особливо міжнародні порівняльні дослідження, є дуже дорогими. В той же час для вирішення поставлених дослідницьких питань часто можна скористатися даними, зібраними раніше фаховими дослідницькими колективами. В той же час не можна уникнути вторинного аналізу у випадку вивчення динаміки в часі, оскільки знання про минуле можна черпати виключно з проведених раніше досліджень. Саме тому зараз вторинному аналізу приділяють велику увагу. У випадку великих проєктів, зокрема міжнародних порівняльних масштабних проєктів, зібрані в ході виконання проєкту дані обов'язково розміщуються у публічно доступних архівах. Доступ до даних в таких архівах здійснюється через Інтернет.

Однією з важливих цілей створення у ряді країн Європи архівів даних соціальних досліджень є надати можливість дослідникам та особливо студентам і аспірантам дані для вторинного аналізу. Як правило, архіви не займаються проведенням досліджень та збиранням даних, а є лише посередниками між тими, хто збирає (а отже "продукує") дані та академічною спільнотою. Можна виділити три основні групи "продуцентів" даних - це некомерційний сектор (держані установи/агенції, недержавні організації тощо), комерційний сектор (фірми, що займаються маркетинговими дослідженнями, опитуваннями громадської думки, експертними опитуваннями на замовлення тощо) та академічні дослідницькі

установи. Архіви не тільки зберігають дані та роблять їх доступними для інших дослідників чи колективів, але й виконують певну важливу роботу, що підвищує цінність даних. Мова йде про те, що:

- перед депонуванням дані в архіві піддають додатковим процедурам перевірки та очищення для забезпечення несуперечливості та цілісності даних;
- з даних видаляються залежності від специфічних програмних засобів та нестандартних форматів, щоб бути впевненим у технічній можливості читати та використовувати депоновані в архіві дані у майбутньому;
- виконується комплекс робіт по каталогізації та інтеграції даних, що депонуються в архіві, із іншими, вже доступними в архіві, даними; це дає змогу, зокрема, включити ці дані до можливих процедур ефективного пошуку;
- для даних створюється стандартна документація на електронних носіях.

Отже, значну частину роботи сучасного архіву складає саме опис даних, тобто створення метаданих. Саме метадані забезпечують ефективний пошук, надають можливість вірно інтерпретувати дані, забезпечують включення даних у "науковий обіг".

Потрібно зазначити, що необхідність у чіткому та по можливості стандартизованому опису результатів експериментів, експедицій, опитувань тощо є характерною для багатьох наук. Іншими словами, метадані потрібні не тільки в сфері емпіричних соціальних досліджень. Але в кожній конкретній сфері є своя специфіка і надалі ми будемо спиратися на досить значний досвід роботи із метаданими саме в сфері емпіричних соціальних досліджень, який взагалі кажучи не обмежується тільки створенням архівів соціальних даних.

Часто визначають коротко, що "метадані це є дані стосовно даних" [1], тобто інформація про інформацію. Тобто це є структуровані дані, що виступають характеристиками певних сутностей (тих сутностей або ж ресурсів, опис яких цікавить дослідника, які представлені в базі даних тощо), та призначені для їх (сутностей) ідентифікації, пошуку, оцінювання, відповідного управління тощо. По суті метадані емпіричного соціального дослідження представляють собою всю сукупність документів, що описують не тільки зібрані в ході дослідження дані але й багато процедурних та контекстних деталей дослідження. В цій сукупності можна виділити, зокрема, наступні блоки (підмножини):

- структурні метадані, що описують структуру даних. Мова йде про перелік змінних, про можливі визначені зв'язки між змінними, про те яка змінна займає яку колонку в таблиці тощо;
- описові метадані, що описують час та спосіб збирання даних, особливості та властивості вибірки, показники якості вимірювання тощо;
- метадані опису управління даними, що фіксують замовника та власника даних, обмеження доступу до даних, правила використання та посилання на дані в публікаціях тощо;
- метадані про поведінку учасників дослідження (наприклад, про поведінку респондентів під час проведення інтерв'ю), що часто називають "параданими".

Важливо зазначити, що метадані призначені не тільки і не стільки для людей скільки для комп'ютерної обробки. Це висуває досить жорсткі вимоги до формалізації та структурування метаданих. Причому кожен "різновид" метаданих

має своє призначення. Зокрема, структурні метадані призначені в першу чергу для організації читання, конкретного сприйняття, перетворення та навіть об'єднання файлів даних у різних форматах. В той же час метадані описові та про поведінку учасників дослідження в першу чергу спрямовані на забезпечення можливості оцінювання якості даних.

Часто критика використання метаданих для організації роботи із даними соціальних досліджень спирається на те, що сама структура метаданих не є чітко визначеною, метадані є залежними від контексту конкретного дослідження, а отже метадані погано піддаються стандартизації. В той же час саме стандартизація метаданих розкриває можливості розробки прийомів пошуку даних не просто в межах певного архіву, але в Інтернеті в цілому, тобто в середовищі наповненому слабо структурованими різнорідними інформаційними об'єктами. В цій частині концепція метаданих досить тісно пов'язана із концепцією семантичного Інтернету. Дискусії про доцільність та ефективність використання метаданих все ще тривають і в той же час паралельно розробляються моделі метаданих та відповідні інструменти, що претендують на роль стандартів, зокрема Data Documentation Initiative [2], Dublin core [3], Statistical Data and Metadata Exchange [4] та інші. Стандарт для представлення, інтерпретації та обробки метаданих є необхідним для того, щоб дозволити створити автоматичну систему управління та обробки метаданих, яка могла б на основі малоформалізованих запитів здійснювати пошук даних, інтеграцію даних з ряду джерел та побудову агрегованих даних, оцінку якості відібраних даних тощо.

Метадані не тільки містять інформацію, що використовується при пошуку даних. Також в метаданих може фіксуватися мета дослідження (включаючи і дослідницькі гіпотези), процедурні (яким чином відбувалося дослідження) та контекстні особливості дослідження. Зокрема мова може йти про час, місце, про події, що відбувалися безпосередньо перед та під час дослідження тощо. Метадані можуть стосуватися як дослідження в цілому, так і окремих складових дослідження. Зокрема, мова може йти про окрему змінну у файлі даних, про блок змінних або ж про спрямований на вивчення певної проблеми дослідницький модуль. При цьому такий опис пов'язаний із цією складовою безпосередньо і таким чином, що вилучення або копіювання цієї складової веде до копіювання і відповідних метаданих. Дані завжди копіюються виключно із відповідними метаданими (це стосується і частин даних). Тільки за таких умов залишається можливість потім аналізувати скопійовані дані, здійснювати об'єднання та агрегацію, коректно використовувати дані, здійснювати пошук у результатах пошуку, здійснювати коректні перетворення одного формату в інший тощо.

Більше того, метадані можуть накопичувати також історію аналізу та історію обчислень. Зокрема, для змінної важливим може бути чи відображає вона первинну (забрану в ході спостережень) інформацію чи є певним чином обчисленою. У другому випадку важливим може бути не тільки яким чином змінна була обчислена, але й те, які характеристики вона отримала в результаті саме такого обчислення. Наприклад, для змінної V_{123} може бути вказано, що вона є обчисленою як середнє значення змінних V_{24} , V_{28} , V_{32} , V_{36} , V_{40} і коефіцієнт надійності альфа Кронбаха для такого адитивного індексу

дорівнює 0,85. Подібна інформація може бути дуже важливою для вторинного аналізу.

Саме метадані надають сенс (інтерпретацію) числовим даним. Іншими словами, дані у соціальних науках стають доступними для аналізу, для роботи лише через метадані. Без відповідного опису (в тому числі і неформалізованого, звичайною мовою) дані часто представляють собою лише послідовність чисел, що майже позбавлена сенсу. Крім того без адекватного, добре структурованого опису дослідження та, можливо, певних його частин неможливим є автоматизований пошук цього дослідження. Дослідження стає незапитаним, не цитується, не використовується в роботах інших фахівців, а отже певною мірою вилучається з наукового обігу.

Концепцію метаданих можна також розглядати як розширення та розвиток концепції так званих "самодокументованих" файлів даних, що стосується не тільки даних емпіричних соціальних досліджень. Метадані містять інформацію про певні ресурси (зокрема, про ресурси банку даних). Метадані можуть "вбудовуватися" безпосередньо у самі ресурси, або ж можуть бути розташованими у окремих документах, які є окремими від ресурсів, що їх описують метадані, та існують паралельно та незалежно від цих ресурсів. Безпосередньо ресурси банку даних призначені для сприйняття (аналізу) або людиною або певною комп'ютерною програмою. В той же час метадані призначені для організації пошуку та автоматичного створення плану аналізу. Метадані дають можливість здійснювати певні логічні виведення (робити логічні висновки) про властивості ресурсів і на основі цих висновків будувати схеми пошуку та/або схеми аналізу.

Техніка гіперпосилань стала настільки звичною користувачам Інтернету, що нас вже не дивує можливість зв'язати фразу у власному тексті із фотокарткою або главою книги, що знаходиться у бібліотеці на іншому континенті. Якщо на початку, коли тільки створювався Інтернет та відповідні протоколи зв'язку, гіперпосилання використовувалися для того, щоб зв'язати між собою два текстових фрагменти або у одному документі або ж у двох різних документах, то зараз гіперпосилання зв'язують між собою будь-які "оцифровані" об'єкти – тексти, малюнки, звуки тощо. Якщо мова йде про емпіричні соціальні дослідження, то є принаймні два різновиди таких цифрових об'єктів, які є сенс зв'язувати такого роду посиланнями - це наукові тексти (звіти, статті, виступи на наукових зібраннях, книги тощо) та емпіричні дані (на яких, власне, і ґрунтуються наукові тексти). Для текстів є чудово відпрацьовані стандарти структурування та посилань, є відповідні програмні засоби (включаючи стандартні браузері, програми перегляду на зразок Acrobat Reader тощо). Крім того тексти є "самовизначеними" або "самодокументованими" об'єктами. Це означає, що зміст об'єкту є зрозумілим із самого цього об'єкту (зміст тексту є зрозумілим після того, як текст є прочитаним). В цьому сенсі тексти не потребують додаткового опису, хоча і для текстів використовують такі метадані, як ключові слова (полегшують пошук та класифікацію), класифікаційні коди (на кшталт УДК), резюме, інформація про авторів тощо. Зазвичай вся ця інформація міститься в самому тексті і може, при необхідності, бути знайдена в ньому спеціальними комп'ютерними програмами (є навіть програми автоматичного реферування

науково-технічних текстів), але явне виділення цієї інформації у вигляді метаданих дає можливість працювати із текстами більш ефективно.

Що ж стосується емпіричних даних, то ситуація у порівнянні із текстами є значно складнішою. Емпіричні дані не є "самовизначеними" або "самодокументованими", їх зміст стає зрозумілим лише за використання відповідних метаданих. Більшість комп'ютерних програм аналізу, що ведуть свій розвиток ще з 1980-х років, не орієнтовані на накопичення та автоматичну обробку метаданих і в кращому випадку містять можливість зберігати метадані призначені не для автоматичної обробки, а для інформування дослідника, що здійснює підготовку та аналіз даних. Зокрема, у пакеті SPSS, що по суті став стандартом де-факто для соціальних дослідників, є можливість зберігати мітки змінних та мітки окремих значень. Ці мітки призначені для інформування дослідника про зміст змінної або конкретного значення, але не призначені для автоматичної обробки. Є можливість надавати специфічну інтерпретацію певним значенням змінної, інтерпретацію "відсутнього значення", що автоматично веде до специфічних правил роботи із такими значеннями. Є можливість вказувати інформацію про шкалу змінної, але можливості використання такої інформації є дуже обмеженими. Натомість у вітчизняному пакеті статистичного аналізу даних ОСА, який створювався у другій половині 1990-х років, поряд із мітками змінних та окремих значень, а також із визначенням окремих значень як кодів "відсутніх значень" є також можливість визначати тип шкали змінної і цей тип шкали активно використовується під час обчислень. Є можливість вільного, орієнтованого на дослідника, коментування як кожної окремої змінної, так і файлу даних в цілому. Крім того пакет автоматично зберігає історію обчислення змінної та навіть певну інформацію про статистичні властивості обчисленої змінної (зокрема показник надійності для адитивного індексу). Автоматизація створення метаданих в процесі роботи із даними є важливою, оскільки тісно пов'язана із людським фактором. Зазвичай люди не хочуть спеціально займатися описом даних та створенням метаданих. Автори дослідження з більшим задоволенням та з більшою користю для власної кар'єри будуть писати публікації, що ґрунтуються на зібраних даних, ніж структурувати та формально описувати ці дані.

В деяких випадках метадані є включеними до самих даних, що може ускладнювати роботу із ними. Як приклад можна назвати наступне посилання на дату проведення опитування у тексті неформалізованого інтерв'ю - "опитування проведено за десять днів до президентських виборів 2010 року". Робота із таким посиланням, яке по суті є метаданими, є ускладненою, особливо якщо вести мову про автоматичну обробку. Тому навіть у випадку, коли відповідні метадані є безпосередньо частиною основного документу, корисним є використання принаймні спеціальних позначок (так званих "тегів"), які значно полегшують виділення метаданих із загального тексту.

Крім того предметом дискусії може бути і правильність (достовірність) інформації у метаданих. Адже ця інформація, як і будь-яка інформація, може бути не повною мірою достовірною. Свідомо або несвідомо до метаданих можуть потрапляти помилки або ж перекручення. Це ставить задачу перевірки цілісності, правильності, несуперечливості метаданих. Іншими словами ми маємо для метаданих виконати той самий обсяг робіт, що і безпосередньо для даних. Створюючи паралельно самій базі даних ще й базу метаданих ми по суті

подвоюємо і обсяг роботи і обсяг проблем. Адже для бази метаданих ми маємо вирішити комплекс тих самих проблем, які ми намагаємося вирішити для самих даних із використанням метаданих як інструменту. Звичайно полегшенням є те, що метадані мають формалізовану жорстку структуру та невеликий (у порівнянні із самими даними, що їх метадані описують) обсяг.

З іншого боку метадані за необхідності частково дублюють дані, які вони описують. Ця робота може видаватися зайвою, бо люди не схильні виконувати ту ж саму роботу двічі (записувати те ж саме кілька разів різними словами, різними формальними мовами, з використанням різних інструментів). В цьому сенсі система створення, ведення та супроводу метаданих має бути необтяжливою. Бажано, щоб вона стала природною частиною самого процесу дослідження, допомагала досліднику (зокрема, прояснювала йому логічну структуру дослідження, взаємозв'язок певних компонентів дослідження тощо), а не виступала зовнішньою та непотрібною самому досліднику. Працювати лише заради того, щоб дослідження стало доступним іншим, витратити на це свій час - це може бути поганою мотивацією для дослідника в плані ведення метаданих. Саме тому автоматизація створення метаданих, або принаймні їх частини, в процесі роботи із даними не тільки значно полегшує роботу з перевірки цілісності та достовірності, але й взагалі сприяє самому накопиченню метаданих.

З появою та розвитком Інтернет розпочалися роботи по створенню процедур (алгоритмів) уніфікованого пошуку та доступу до даних саме у середовищі всесвітньої мережі. Така діяльність потребує за необхідності високого рівня стандартизації не тільки та не стільки представлення даних, скільки стандартизації представлення метаданих, оскільки пошук даних та доступ до них (контроль доступу, визначення прав доступу, облік кількості звернень, тощо) здійснюється саме на основі метаданих. При цьому у проектуванні відповідних стандартів за модель представлення зібрання даних беруть не архів, а бібліотеку, що має у порівнянні із архівом більш прості, більш публічні умови доступу до інформації. Структура метаданих виступає як гнучка, що може із часом розвиватися та змінюватися. За таких умов метадані дають можливість зв'язати із набором даних (колекцією наборів даних) відповідні публікації, навчальні матеріали, описи історії та перспектив або планів аналізу тощо. Іншими словами такий підхід дозволяє створювати навколо даних поступово, крок за кроком, певне "інтелектуальне середовище". Такий розвиток є більш реалістичний, ніж одноразове створення відповідного середовища за визначеним планом. Метою ж такої діяльності із створення стандартів є не тільки розробка регламентуючих документів, але й створення публічно доступних програмних інструментів для організації процесів створення та розвитку середовища метаданих, для організації пошуку на основі метаданих, для організації коректного доступу до метаданих з метою вирішення різних задач по їх обробці.

В світі є кілька ініціатив із створення стандартів метаданих та комплексу відповідних документів. У сфері роботи із даними емпіричних соціальних досліджень безперечно одним з найбільш активних є відкритий консорціум DDI (Data Documentation Initiative) [3] - міжнародний проект, що має на меті створення стандарту представлення інформації із описом даних у соціальних науках (міжнародний XML стандарт для метаданих у соціальних науках). Проект розпочався у 1995 р. Інструментом для опису стандарту є мова XML. Версія 3.1

стандарту була оприлюднена у 2009 р. Стандарт фіксує форми для представлення, обміну та збереження інформації та спрямований на підтримку повного життєвого циклу даних соціального дослідження. Метою є створення таких структур метаданих, які б повністю були придатні для комп'ютерної обробки.

Важливим є тісний зв'язок між метаданими та відповідними даними. Досить розповсюдженою є ситуація, коли метадані зберігаються окремо від самих даних. Для роботи із такими метаданими, для пошуку із використанням метаданих використовується специфічне програмне забезпечення, відмінне від програмного забезпечення роботи із самими даними. Це пов'язано із тим, що, як зазначалося раніше, стандартні, найбільш розповсюджені та часто вживані пакети програм аналізу даних не містять можливостей не тільки працювати, але й зберігати метадані (формати даних не передбачають такого), а отже для зберігання метаданих та роботи із ними розробляють відповідне специфічне програмне забезпечення. Тому важливою діяльністю консорціумів із розробки стандартів метаданих є створення інтегрованих спеціалізованих програмних засобів, що призначені для заповнення розриву між роботою із метаданими (пошук у метаданих, аналіз метаданих) та аналізом даних, що їх метадані описують. Прикладом може бути система NESSTAR [5], що дозволяє створювати метадані для опису емпіричних даних, публікувати в Інтернет описані таким чином дані та надавати доступ з метою статистичного аналізу до цих даних. Це проект значною мірою спирається на стандарт DDI. Більше того, тісний зв'язок метаданих та даних сприяє поліпшенню (полегшенню) інтерпретації результатів аналізу даних, оскільки опис саме тих частин даних, що аналізуються, є корисним для розуміння (інтерпретації) результатів статистичного кількісного аналізу.

Ідеальне (або ж близьке до такого) середовище для соціального дослідника, що активно використовує емпіричні дані, можна було б визначити наступним чином.

1. Всі емпіричні дані (статистика з різних джерел, дані емпіричних досліджень тощо) розміщені у мережі і є доступними on-line.

2. Існує та є доступною уніфікована універсальна пошукова система для роботи із ресурсами, через які можливим є доступ до даних (із архівами, банками та базами даних чи навіть зібраннями файлів досліджень окремих дослідницьких колективів).

3. Із емпіричними даними та/або їх частинами пов'язаний значний обсяг метаданих, що є основою для ефективного пошуку, виділення частин даних, конвертації даних, аналізу даних тощо

4. Дослідник має можливість виділяти потрібні частини у знайдених даних та копіювати їх з мережі у потрібному форматі на свій локальний комп'ютер для подальшої роботи.

5. Спеціальні програми ("помічники дослідника", так звані knowbots [6]) відслідковують появу у мережі нових даних, що можуть зацікавити дослідника (у відповідності із визначеною та описаною засобами опису метаданих сферою інтересів дослідника), та інформують про це дослідника.

6. Метадані, пов'язані із даними емпіричного дослідження, містять посилання на всі публікації (звіти, статті, книжки, виступи на радіо або

телебаченні тощо), що базуються на цих даних, та посилання на відповідних дослідників та/або дослідницькі колективи.

7. За необхідності дослідник, що звернувся до даних, може самостійно додати власну інформацію до метаданих.

Ситуація, що існує зараз із використанням метаданих, не відповідає сучасним інформаційним технологічним можливостям. Далеко не всі дані є публічно доступними. Різні дослідницькі організації та навіть державні агенції мають різні правила публікації даних свої досліджень. В деяких випадках дані стають публічно доступними лише через певний час після завершення дослідження. В деяких випадках публікуються лише звіти в той час як безпосередні дані не виставляються в публічний доступ. Чи не єдиним виключенням із цих "традицій" є масштабний міжнародний порівняльний проект, що має назву "Європейське соціологічне дослідження" (ESS, European Social Survey) [7]. Дані цього проекту стають вільно публічно доступними в Інтернет одразу після завершення польового етапу опитування та депонування даних у Норвезькому архіві соціальних даних. Крім того досі немає єдиного стандарту метаданих, а отже відсутньою є уніфікована система пошуку та доступу до тих даних, що є цікавими та потенційно придатними до аналізу. Проблему також складають досить розповсюджене небажання дослідників та окремих колективів робити власні дані загальнодоступними (наукова конкуренція), певні юридичні (права інтелектуальної власності на дані) та комерційні (інформація має вартість і часто досить значну) перешкоди. Все це збільшує дистанцію між дослідником та потрібними йому ресурсами із статистичними даними. А якщо взяти до уваги ту вже загальновизнану думку, що справжня цінність даних відкривається (а також, на відміну від більшості матеріальних цінностей, зростає) через активне їх використання, то існуюча на даний момент ситуація не є задовільною.

Висновки. Використання стандарту метаданих надає можливість ефективно обмінюватися даними, поширювати дані за межі колективу, що створив ці дані, а це є важливим, зокрема, як для моніторингових так і взагалі для порівняльних досліджень. При цьому дослідник отримує інструмент, що дозволяє не тільки описувати дані соціального дослідження із максимально можливою повнотою (тобто створювати відповідні метадані), але й допомагає без значних та зайвих зусиль підтримувати відповідні метадані у актуальному стані. Якісні метадані не тільки забезпечують релевантний пошук даних, але й можуть бути використані для автоматичного планування аналізу даних. Важливим також є те, що оснащені якісними добре структурованими та стандартизованими метаданими результати емпіричного дослідження не обов'язково мають знаходитися в певно визначеному місці, наприклад у електронному архіві, а можуть бути розміщені в Інтернет на сторінці дослідницької групи або навіть на персональній сторінці окремого дослідника. Метадані у сукупності із інструментами пошуку та доступу, орієнтованими на відповідний стандарт, забезпечують включення таким чином розташованих даних дослідження у науковий обіг такою ж мірою ефективно, як і у випадку розміщення даних у публічних електронних архівах.

Стандартизація представлення текстів у електронному форматі у сукупності із інструментами пошуку та перегляду призвела до того, що досить часто невеликі електронні бібліотеки чи навіть приватні зібрання електронних текстів

конкурують за популярністю із великими бібліотеками. Так само стандартизація метаданих у поєднанні із децентралізацією зберігання даних емпіричних досліджень може з одного боку зменшити відстань між тими, хто продукує дані, і тими, хто їх використовує, а з іншого боку може призвести до різкого зростання в Інтернет кількості оцифрованих та потенційно придатних до аналізу емпіричних даних, що до певної міри може створити конкуренцію існуючим архівам соціальним даних.

Література

1. D.Marco, M.Jennings Universal Meta Data Models.- Wiley Publishing.- 456 pp.
2. <http://www.ddialliance.org/what>
3. <http://dublincore.org/>
4. <http://sdmx.org/>
5. <http://www.nesstar.com>
6. J.Ryssevik S.Musgrave The Social Science Dream Machine: Resource Discovery, Analysis, and Delivery on the Web // Social Science Computer Review, Vol. 19 No. 2, Summer 2001 p. 163-174
7. <http://www.europeansocialsurvey.org>